Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5828129, 7 pages
https://doi.org/10.1155/2021/5828129

Research Article

Application of Flipped Classroom Model Driven by Big Data and Neural Network in Oral English Teaching

Yujun Zeng

College of Foreign Languages of Hunan Institute of Engineering, Xiangtan 411101, China

Correspondence should be addressed to Yujun Zeng; 05012@hnie.edu.cn

Received 10 July 2021; Accepted 12 August 2021; Published 26 August 2021

Academic Editor: Yuanpeng Zhang

With the advancement of big data and neural network technology, flipped classroom informatization has shifted the traditional order of knowledge transfer and internalization, emphasizing students' autonomous learning before class, knowledge absorption, and knowledge completion in class with the assistance of teachers. Students' internalization and consolidation create the conditions for individualized learning. In foreign teaching, the benefits and feasibility of the flipped classroom have been demonstrated, and it is a promising new teaching model. Although recent research on oral English teaching in Chinese universities has yielded promising results, students' classroom activity and participation remain low, learning initiative is lacking, and opportunities and time for oral training are insufficient. This article uses flipped classroom, big data, and neural network technology to teach college oral English classes, with the goal of determining whether the flipped classroom model can help students improve their oral English proficiency and self-learning ability, as well as exploring students' attitudes toward the flipped classroom model. This paper first proposes a big data and deep neural network-based algorithm for detecting oral English pronunciation errors, which can be used for self-correction of students in the flipped classroom mode to improve the quality of oral English teaching. Finally, we also conducted simulation experiments, and the experimental results show that our algorithm is 4.12% better than SVM.

1. Introduction

Although a series of teaching reforms [1–3] have been carried out, there are still many problems in English teaching in China [4–6]. For example, students lack the ability of independent learning [7–9] and cooperative learning [10–12]. The teacher's lack of professional quality and the solidification of educational concept are the unfavorable factors that affect the teaching effect. "Flipped classroom" teaching mode [13–16] is a new form of teaching and classroom organization. Many new inspirations and concepts will be integrated and put into practice in the teaching process of flipped classroom. It not only stimulates students' interest in learning, but also develops students' independent learning ability and cooperative research ability. Flipped classroom subverts the traditional classroom model [17] and also changes the role of teachers and students in traditional teaching. The time in the classroom is rearranged, and the time between knowledge transfer in class and internalization of knowledge absorption after class is adjusted, resulting in traditional classroom innovation.

Big data [18] and neural network technology [19–21] have transformed how people communicate. People are constantly learning new things and interacting with others. Technology has had a significant impact on not only how people work and live, but also on how they learn. Digital technology is used to record lecture content, as well as audio and video related to teaching knowledge, providing technical support for flipped classroom implementation. Furthermore, after class, students must self-study online knowledge content, freeing up valuable classroom time to address students' questions and difficulties or engage in other constructive classroom activities. Instead of doing what the teacher says in class and doing homework and exercises after class, you should do what the teacher says. The teaching method is flipped, with students receiving explicit instructions to learn after class and then applying what they have learned in class.

The most significant advantage of flipped classroom is that it can maximize the effect of classroom learning [22], which is also the most basic demand of teaching. In the liberated classroom time, the teacher can give personalized guidance to each learner according to his/her characteristics and knowledge mastery, so as to make the classroom time more meaningful. Moreover, more flexible learning tasks and learning pace can be provided in the schedule according to each student's learning progress. Educators have benefited from technological advancements in education for a long time by incorporating technology into their teaching. Students can access online learning at any time and from any location thanks to popular online services like online video sites and Baidu Cloud. Students can easily follow a variety of online courses by clicking on links on social media, which is very convenient and flexible. As a result, a lot of flipped classroom big data is generated.

Based on the foregoing observations, we can conclude that, while recent research on oral English teaching [23] in Chinese universities has yielded promising results, students' activity and participation in the classroom remain low, learning initiative is lacking, and opportunities and time for oral training are insufficient. This article explores students' attitudes toward flipped classroom teaching mode by applying flipped classroom, big data, and neural networks [24–26] to college oral English classroom teaching, which helps to improve students' oral English level and autonomous learning ability.

The main contributions of this paper are as follows:

(1) This paper applies flipped classroom, big data, and neural network technology to college oral English classroom teaching, which helps to improve students' oral English level and autonomous learning ability, and explores students' attitudes toward flipped classroom teaching mode

(2) This paper proposes a pronunciation error detection algorithm based on phonetics space and KL divergence distance metric. The basic element of the phonetics space is senone, and its posterior probability is obtained by discriminative learning from the acoustic features by the deep neural network

The rest of the paper is organized as follows. In Section 2, the background is outlined. The methodology is conducted in Section 3. The experimental results are further summarized in Section 4. Finally, Section 5 concludes the paper with a summary and future research directions.

2. Background

At present, emerging information technology [27, 28] provides a powerful driving force for economic and social change and development. Facing the knowledge explosion in the context of the new era and the rapid changes in technology, learners have new learning needs. In the information age, how to cultivate people also puts forward new requirements. Make full use of information technology and strive to create an information-based learning environment for learners, so as to promote the reform of teaching concepts, teaching models, and teaching content. In the daily teaching process, information technology is continuously deepened and widely used, so as to better serve the new talent training services under the background of the times. In the context of the new era, the intelligentization of education should play an active leading role, and the intelligentization of education should support and lead the modernization of education. Simultaneously, education informatization is critical to the renewal of educational concepts, the reform of teaching models, and the reconstruction of the educational system. The incorporation of big data and neural network technology into the process of subject teaching innovation, as well as the realization of deep integration between information technology and subject teaching, has become a hot topic of current research in this context. Teaching concepts and teaching activity design continue to innovate and present a diverse development as a result of continuous advancement and in-depth research.

In terms of classroom form, teaching process, and teacher and student roles, the flipped classroom has been flipped. The format has shifted from "teaching first, then learning" to "first, then teaching." The teaching process is being turned on its head thanks to information technology. Teachers are now "teaching," and the traditional role of students as "learners" has shifted. Flipped classrooms overcome the limitations of traditional teaching. Teaching places a greater emphasis on students' learning processes, allowing them to have a more authentic learning experience. In the teaching process, more attention is paid to student self-study, teacher-student interaction, and teacher-student, student-student interaction. Collaborative exchanges between students have changed the former difficulty of passively accepting knowledge and help students develop their lifelong learning capabilities. From its emergence to its continuous development, flipped classrooms have been sought after by different scholars and educators all over the world and have had a profound impact. Objectively speaking, the flipped classroom has improved the teaching quality of the subject and the ability of students, and it has had an important reform effect on classroom teaching. However, flipped classrooms have not set off a wave of reform in China. Aside from the higher demands of flipped classrooms in a networked teaching environment, teachers' educational concepts and teaching viewpoints must be updated urgently. Nowadays, education is becoming increasingly informatized, and educational infrastructure is being built to a higher standard. To promote the depth between information technology and subject teaching, teachers should learn to use networked learning platforms to make full use of advanced teaching concepts such as blended learning and flipped classrooms.

3. Methodology

3.1. Flipped Classroom. Flipped classroom is a teaching mode that mixes direct explanation and constructivist learning. The traditional teaching model is mainly divided into three parts: preclass knowledge preview, in-class knowledge teaching, and after-class knowledge consolidation. The "flipped classroom" breaks the traditional teaching model and reverses the process of knowledge transfer and internalization, that is, teachers send relevant learning materials to WeChat, QQ, and other online platforms before class, and students complete the teaching content independently before class. Classroom time is mainly used to carry out various teaching activities, such as teachers explaining the problems existing in students' self-learning before class and students exploring problems in groups, so as to obtain a deeper level of new knowledge. Understand, digest, and absorb. Flipped classroom extends the classroom from school to home, greatly enhancing the openness of education. Students can formulate their own learning 4 progress according to their own learning ability and will no longer be affected by the uniform teaching of teachers. This gives full play to the autonomy and initiative of students in learning, and individualized learning is also guaranteed, so as to achieve a better educational effect. The flipped classroom teaching model is shown in Figure 1.

3.2. Flipped Classroom Teaching Model of Spoken English

3.2.1. Basic Model. The timeline of a flipped classroom generally includes two parts: before class and during class. Before class, it is used for students to study independently, that is, the stage of "knowledge externalization." The class is used for the development of learning activities, that is, the stage of "knowledge internalization." Due to the differences in learners, implementation environment, learning methods, and other subjects in oral English teaching, the author takes Professor Robert Tallbert's flipped classroom implementation structure model as the basis and combines the teaching content and teaching objects of this flipped classroom. Characteristic, the structure model of the flipped classroom implementation of Professor Robert Tallbert has been modified and refined accordingly, and the teaching implementation model of the flipped classroom of spoken English has been constructed. The specific model is shown in Figure 2.

The textualization of the flipped classroom teaching implementation model in Figure 2 can be divided into the following four steps in detail:

(1) *Teachers Provide Learning Materials.* There are many types of preclass learning materials for flipped classrooms, such as teaching microvideos, presentations, and teaching cards. Teachers can choose according to teaching objects, teaching content, and actual teaching needs. But generally speaking, the preclass learning materials of the flipped classroom are mainly teaching microvideos, and teachers can download or make their own teaching resources from related websites. Famous foreign teaching video websites include Khan Academy and TED-ED. Teachers can look for video resources that match their teaching content among high-quality open educational resources as the course teaching content. After the teachers prepare the learning materials, they can upload them to the learning platform.

(2) *Students' Autonomous Learning Stage before Class.* First, students learn the learning materials provided by the teacher before class to complete the task of autonomous learning. If students encounter problems in the learning process, they can communicate and discuss with their classmates on the platform to find solutions. Problems that cannot be solved can be fed back to the teacher individually, and the teacher will give a one-to-one personalized explanation on the platform. Secondly, students independently complete the test questions uploaded by the teacher to the platform, and the teacher roughly determines the student's preclass learning effect based on the students' test situation. Finally, students will perform preclass exercises according to the oral tasks assigned in the self-learning task list, paving the way for oral English communication in the classroom.

(3) *In-Class Knowledge Internalization Stage.* Before class, students have already learned the relevant teaching knowledge points of this class. This provides ample time for the internalization of knowledge in the middle of the class. First, the teacher will give a unified explanation of the questions with a higher error rate according to the students' test situation before class and explain the relevant knowledge points from the wrong questions again. Second, conduct classroom testing. The methods of classroom testing can be diversified, such as teachers asking questions, students answering, or students asking and answering each other in a game. Third, let students discuss the knowledge expansion questions in the learning task list in groups and share the results of the discussion after the end. The teacher comments and supplements the results of the students' discussion. Then, carry out classroom activities. As it is an oral class, teachers can set up some classroom activities aimed at improving students' oral communication skills. For example, students are grouped to carry out activities such as situational dialogue, role playing, and microvideo dubbing. In the course of the activity, teachers should give individual tutoring to individual student with poor oral English, encourage them to speak English, actively participate in classroom activities, and encourage students to take the initiative to show their results after the end.

(4) *Evaluation and Summary Stage.* Let students conduct self-evaluation and mutual evaluation on their own learning attitude, self-confidence, cooperative spirit, and spoken English expression in this lesson, cooperate in groups to summarize the key and difficult points of this lesson, and complete the unit review mind map. After the end, the group representatives went to the stage to show the unit mind map and explained the important and difficult points of this lesson, the teacher made comments and supplements, and the students made corresponding notes.

3.3. Spoken Pronunciation Detection Algorithm. Figure 3 depicts the structure of the proposed pronunciation error detection model in this paper. It is divided into two sections:

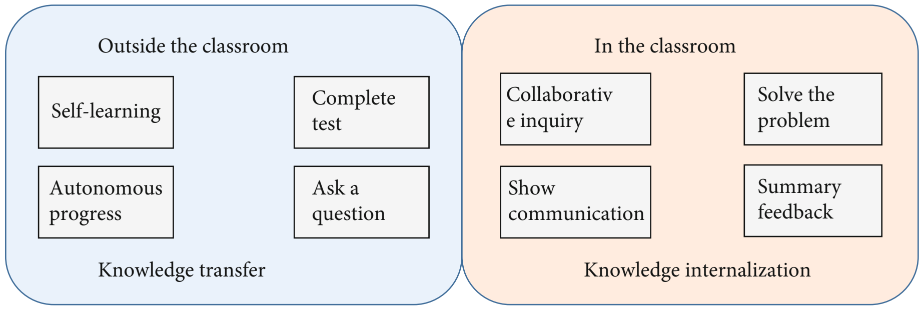

Figure 1: The teaching mode of flipped classroom.

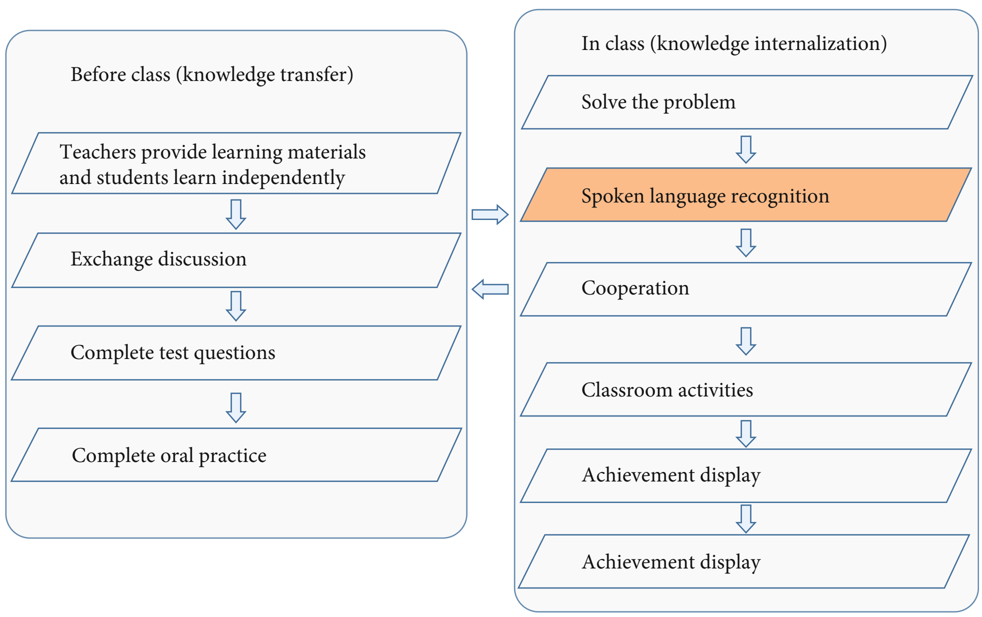

Figure 2: An implementation model of flipped classroom teaching of spoken English.

input characteristics and classifier design. The two parts will be discussed in greater depth below.

3.3.1. Input Features. In view of the fact that acoustic features, such as MFCC and PLP, are susceptible to noise interference from different speakers, transmission channels, and recording environment, this paper adopts high-dimensional phonetics features as the basic features of pronunciation error detection. Suppose there are M different senses $\{s_1, s_2, \cdots, s_M\}$ in the acoustic model, for any speech frame x_t, its characteristics in phonetics space are expressed as follows:

$$f(x_t) = [p(s_1|x_t), \cdots, p(s_M|x_t)]^T, \quad (1)$$

where the posterior rate $p(s_i|x_t)$ is the output value of the ith node of DNN.

Suppose that the phoneme of speech segment X is b, and three hidden states bs_1, bs_2, bs_3 are obtained through DNN. For the speech segment corresponding to each state, the center vector is calculated to indicate its position in the speech space. The calculation equation is as follows:

$$\bar{f}(X) = [\bar{p}(s_1|X), \cdots, \bar{p}(s_i|X), \cdots, \bar{p}(s_M|X)]^T, \quad (2)$$

$$\bar{p}(s_i|X) = \frac{1}{t_{ek} - t_{sk} + 1} \sum_{t=1}^{t_{ek}} p(s_i|X_t). \quad (3)$$

If the total number of sensors in the acoustic model is M, and each contains 3 valid states, the dimension of the feature vector is 3M. Feature direction $f(X)$ will be used as the input feature of the phoneme classifier for classifier training.

3.3.2. Classifier. The task of detecting pronunciation errors for each phoneme is abstracted as a binary classification problem. Assuming that there are M different phonemes in the phoneme set, the traditional classification method divides the entire data set into M parts according to the corresponding phonemes and learns a binary classifier for each phoneme separately.

Assume that the training data set is $S = \{X_n, t_n\}_{n=1}^N$, which contains M different phonemes $\{p_1, p_2, \cdots, p_M\}$, the phoneme label corresponding to the nth sample is $C_n \in \{1, 2, \cdots M\}$, and its pronunciation is correct or wrong is $t_n \in \{0, 1\}$, where 1 means correct pronunciation and 0 means wrong

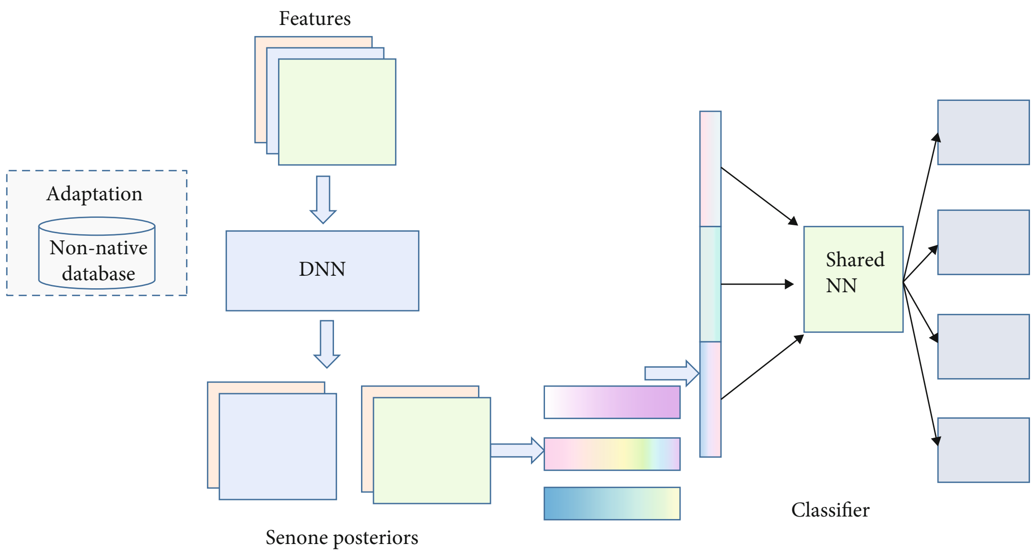

Figure 3: Pronunciation error detection model based on big data and deep neural network.

pronunciation. The maximum likelihood criterion is used to estimate the model parameters, and the optimization objective function is defined as follows:

$$\text{Loss} = \min\,(-\log L),$$
$$L = \prod_{n=1}^{N}\sum_{m=1}^{M}\delta(C_n = m)y_m(X_n)^{t_n}(1-y_m)(X_n)^{1-t_n}, \tag{4}$$
$$y_m(X_n) = \sigma_m\left(W_m^T\phi(X_n)\right).$$

In the testing phase, for sample X, if its corresponding phoneme is marked as i, then its category is judged as follows:

$$p(C|\phi(X)) = \sigma_i\left(W_i^T\phi(X)\right)\begin{cases} \geq\theta & \text{true} \\ <\theta & \text{false} \end{cases}, \tag{5}$$

where θ is the system threshold.

4. Experiments and Results

4.1. Experimental Environment and Parameters. The programming language used is Python, the version is 3.6.5, the deep learning framework used is Keras2.1.5, and the IDE for program deployment is Pycharm, and all experiments are conducted in the same environment. All our experiments have been conducted on a desktop PC with an Intel Core i9-9900K processor and an NVIDIA GeForce GTX 2080ti GPU. We have implemented the model construction through the Keras deep learning library, the programming language we use is Python, and we batch processed 256 samples each time.

4.2. Data Sets. The standard pronunciation database is an English isolated word recognition database named LDC95S27, in which each sentence contains only one English word. The entire data set contains 93,667 sentences, with a total duration of 23 hours. We divide it into two parts: training set and test set. Among them, the training set contains 900 speakers, 6,700+ words in different texts, and the total duration is 20 hours: the test set contains 80 speakers, and the total duration is about 3 hours. The English learning database is collected and recorded by the author. The database contains 60 Chinese people, each of whom reads 300 words, and the word text is randomly generated from LDC95S27. The phoneme-level misreading, interruption, and omission errors of each word are marked by a phonetic expert. The distribution of various pronunciation errors on the entire data set is shown in Table 1. It can be seen from the distribution table that misreading errors account for 77.9% of all pronunciation errors, which is the main type of error.

Table 1: The distribution of the number of various pronunciation errors in the English word learning database.

Type	True	Substitution	Deletion	Insertion
Count	103522	15673	3180	1255

Table 2: The accuracy of pronunciation error recognition of different structure classifiers on the English test set.

Method	200 nodes	400 nodes	600 nodes	1200 nodes
1 layer	93.1	93.4	93.9	92.1
2 layers	93.2	92.9	93.3	92.2
3 layers	91.3	91.9	93.0	92.9

4.3. Evaluation Index. We scored by calculating the distance D_1 and D_2 between the correct pronunciation model C and the wrong pronunciation model IC of the sample center X and the phoneme b. The final pronunciation accuracy score at the phoneme level is expressed as follows:

$$\text{Score} = \frac{e^{-D_1}}{e^{-D_1} + e^{-D_2}}, \tag{6}$$

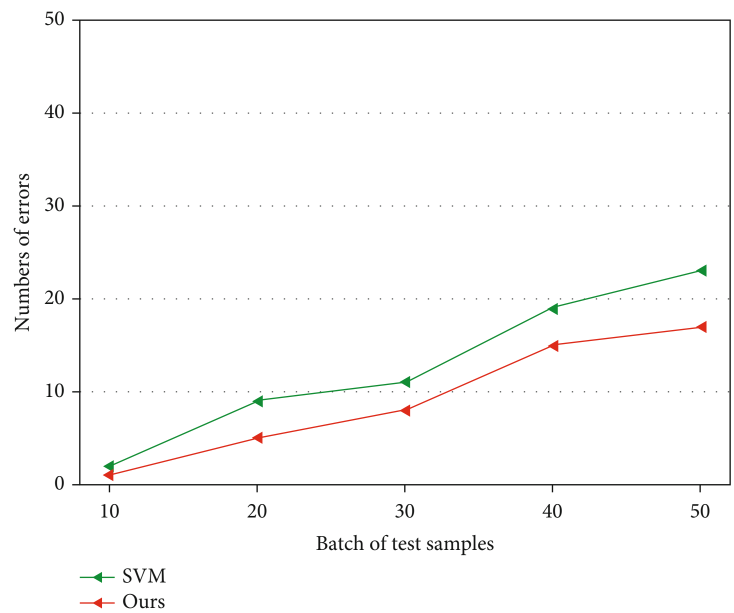

FIGURE 4: Confusion matrix.

$$\text{Score} \geq \theta \begin{cases} T & \text{true} \\ F & \text{false} \end{cases}. \quad (7)$$

4.4. Experimental Results. Firstly, the effects of different hidden layer network structures on experimental results are compared. Because the English learning database is relatively small, the underlying network structure used in the W experiment is also relatively small. The dropout algorithm is used in the training process to prevent overfitting. Table 2 shows the accuracy of pronunciation error detection on the verification set when the number of hidden layers is 1 layer, 2 layers, and 3 layers and the number of nodes in each layer is 200, 400, 600, and 1200. The system threshold is set to 0.4. As can be seen from the table, the artful response of different underlying network structures to the result of articulatory error detection is very small, and the change of accuracy is less than 1%. Among them, the result of the single hidden layer classifier is slightly better than the result of the multihidden layer classifier, which we believe may be caused by the relative shortage of training data. In view of this, the underlying network in subsequent experiments was all set as a single hidden layer, with a number of 600 nodes.

4.5. Comparative Experiment. Support Vector Machine (SVM) is usually used as a classifier in error detection algorithms based on dichotomy. In this section, we compare the performance of SVM and the model presented in this paper in terms of pronunciation error recognition on English learning databases. There are 40 phonemes in the English learning database, so 40 SVM classifiers need to be trained. All SVMs are trained with the SVM-Light tool, and the default parameters are set directly. For each phoneme in a sentence, the classifier predicts the conditional probability that it is pronounced correctly. By adjusting the system threshold value, the number of error recognition under different operation points can be obtained. The pronunciation error recognition results of the two methods on the English learning test set are shown in Figure 4. It can be seen that the model in this paper is overall better than the SVM method, which proves the effectiveness of our model.

5. Conclusion

This paper uses flipped classroom, big data, and neural network technology to teach college oral English classes, with the goal of determining whether the flipped classroom model is effective in improving students' oral English proficiency in Chinese college oral English classes. Students' attitudes toward the flipped classroom teaching model are investigated using oral English ability and self-study ability. This article first proposes an algorithm for detecting oral English pronunciation errors based on big data and deep neural network technology, which can be used to help students self-correct in a flipped classroom setting and improve the quality of oral English instruction. Finally, we ran simulation tests, and the results show that the proposed method outperforms SVM and achieves comparable results.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author does not have any possible conflicts of interest.

Acknowledgments

The study was supported by the Social Science Foundation of Hunan Province (Grant No. 20WLH21).

References

[1] A. Boime, "The teaching reforms of 1863 and the origins of modernism in France," *Art Quarterly*, vol. 1, no. 1, pp. 1–39, 1977.

[2] A. Howes, "Teaching reforms and the impact of paid adult support on participation and learning in mainstream schools," *Support for Learning*, vol. 18, no. 4, pp. 147–153, 2003.

[3] K. Patwardhan, "How practical are the "teaching reforms" without "curricular reforms"?," *Journal of Ayurveda and integrative medicine*, vol. 1, no. 3, pp. 174–176, 2010.

[4] J. R. Cowan, R. L. Light, B. E. Mathews, and G. R. Tucker, "English teaching in China: a recent survey," *TESOL Quarterly*, vol. 13, no. 4, pp. 465–482, 1979.

[5] W. Yi'An, "English language teaching in China: trends and challenges," *TESOL Quarterly*, vol. 35, no. 1, pp. 191–194, 2001.

[6] G. Hu, "English language teaching in China: regional differences and contributing factors," *Journal of Multilingual and Multicultural Development*, vol. 24, no. 4, pp. 290–318, 2003.

[7] S. Y. Hsu, "Building language-learning environments to help technological university students develop English independent learning," *The Jalt Call Journal*, vol. 1, no. 2, pp. 51–66, 2005.

[8] M. N. Islam, "Independent English learning through the Internet," *Journal of Language Teaching & Research*, vol. 2, no. 5, 2011.

[9] N. Pekel, *Students' Attitudes towards Web-Based Independent Learning at Bilkent University School of English Language [Doctoral dissertation]*, Bilkent University, 2002.

[10] M. T. Munir, S. Baroutian, B. R. Young, and S. Carter, "Flipped classroom with cooperative learning as a cornerstone," *Education for Chemical Engineers*, vol. 23, pp. 25–33, 2018.

[11] F. Lestari, B. Saryantono, M. Syazali et al., "Cooperative learning application with the method of" network tree concept map": based on Japanese learning system approach," *Journal for the Education of Gifted Young Scientists*, vol. 7, no. 1, pp. 15–32, 2019.

[12] B. Korbar, D. Tran, and L. Torresani, "Cooperative learning of audio and video models from self-supervised synchronization," 2018, https://arxiv.org/abs/1807.00230/.

[13] J. S. Chen Hsieh, W. C. V. Wu, and M. W. Marek, "Using the flipped classroom to enhance EFL learning," *Computer Assisted Language Learning*, vol. 30, no. 1-2, pp. 1–21, 2017.

[14] B. Ayçiçek and T. Yanpar Yelken, "The effect of flipped classroom model on students' classroom engagement in teaching English," *International Journal of Instruction*, vol. 11, no. 2, pp. 385–398, 2018.

[15] Z. Turan and B. Akdag-Cimen, "Flipped classroom in English language teaching: a systematic review," *Computer Assisted Language Learning*, vol. 33, no. 5-6, pp. 590–606, 2020.

[16] M. Y. Abdullah, S. Hussin, and K. Ismail, "Implementation of flipped classroom model and its effectiveness on English speaking performance," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 9, 2019.

[17] B. M. Smart and D. P. Saxon, "Online versus traditional classroom instruction: an examination of developmental English courses at an Alabama community college," *Community College Journal of Research and Practice*, vol. 40, no. 5, pp. 394–400, 2016.

[18] Z. Wang, P. Zhang, W. Sun, and D. Li, "Application of data dimension reduction method in high-dimensional data based on single-cell 3D genomic contact data," *ASP Transactions on Computers*, vol. 1, no. 2, pp. 1–6, 2021.

[19] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.

[20] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.

[21] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology,*, p. 1, 2021, In press.

[22] P. Zheng, X. Wang, and J. Li, "Exploration and practice of curriculum ideological and Political Construction Reform——-take" information security" course as an example," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 1–5, 2021.

[23] L. Liang, Q. Yin, and C. Shi, "Exploring proper names online and its application in English teaching in university," *ASP Transactions on Computers*, vol. 1, no. 1, pp. 24–29, 2021.

[24] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.

[25] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of ambient intelligence and humanized computing*, vol. 12, no. 8, pp. 8427–8440, 2021.

[26] Y. Zhang, W. Li, L. Zhang, X. Ning, L. Sun, and Y. Lu, "AGCNN: adaptive gabor convolutional neural networks with receptive fields for vein biometric recognition," *Concurrency and Computation: Practice and Experience*, 2020, In press.

[27] J. Zhang, X. Jin, J. Sun, J. Wang, and K. Li, "Dual model learning combined with multiple feature selection for accurate visual tracking," *IEEE Access*, vol. 7, pp. 43956–43969, 2019.

[28] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person reidentification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia.*, p. 1, 2021, In press.